

Multi-criteria analysis using latent class cluster ranking: An investigation into corporate resiliency

Jamshed Mistry^a, Joseph Sarkis^{b,*}, Dileep G. Dhavale^c

^a Faculty of Business Administration, Bilkent University, Ankara, Turkey

^b School of Business, Worcester Polytechnic Institute, 100 Institute Road, Worcester, MA 01609, United States

^c Graduate School of Management, Clark University, United States

ARTICLE INFO

Article history:

Received 8 November 2012

Accepted 6 October 2013

Available online 14 October 2013

Keywords:

Multiple criteria decision making

Performance measurement

Latent class model

Gibbs sampler

Monte Carlo simulation

E-business

Balanced scorecard

ABSTRACT

In this paper, we introduce a multi-stage multiple criteria latent class model within a Bayesian framework that can be used to evaluate and rank-order objects based on multiple performance criteria. The latent variable extraction in our methodology relies on Bayesian analysis and Monte Carlo simulation, which uses a Gibbs sampler. Ranking of clusters of objects is completed using the extracted latent variables. We apply the methodology to evaluate the resiliency of e-commerce companies using balanced scorecard performance dimensions. Cross-validation of the latent class model confirms a superior fit for classifying the e-commerce companies. Specifically, using the methodology we determine the ability of different perspectives of the balanced scorecard method to predict the continued viability and eventual survival of e-commerce companies. The novel methodology may also be useful for performance evaluation and decision making in other contexts. In general, this methodology is useful where a ranking of elements within a set, based on multiple objectives, is desired. A significant advantage of this methodology is that it develops weighting scheme for the multiple objective based on intrinsic characteristics of the set with minimal subjective input from decision makers.

© 2013 Published by Elsevier B.V.

1. Introduction

Performance measurement within organizations typically relies on analyses of multiple factors (Clivillé et al., 2007). The use of multiple factors and relative performance of units are critical for managerial planning and decision-making at both operational and strategic performance levels (Suwignjo et al., 2000; Bititci et al., 2001; Sarkis, 2003). A variety of multiple-criteria, quantitative approaches, and techniques have been developed over the years to address performance analysis issues within increasingly complex organizations. These techniques run the gamut of multiple-criteria decision modeling approaches and frameworks including the balanced scorecard method (BSC) (Kaplan and Norton, 1992); the analytical hierarchy process (AHP), data envelopment analysis, outranking, simple scoring, and numerous other techniques (Koksalan et al., 2011). Each methodology has its own strengths and weaknesses. Many factors such as amount of data required, theoretical robustness, transparency of the approach, level of acceptance by management, amount of time necessary for a solution affect the eventual decision whether or not to adopt a multi-criteria technique.

Given the rich history of the linkage between multiple-criteria analysis and performance measurement, we introduce a novel and relatively robust technique for multiple-criteria evaluation technique based on Bayesian and latent class analysis. The technique is advantageous since it requires very little input from management decision makers, a limited set of data, and has strong theoretical foundations in Bayesian statistical analysis. There are, however, a few disadvantages, which will also be discussed in this paper.

We utilize real world data and an established, well-known managerial technique to demonstrate the usefulness, validity, and flexibility of our methodology. This case example is on the performance evaluation of e-commerce or e-business companies. We utilize the BSC factors in the case example for a number of these organizations. We apply the methodology developed in this paper to rank these e-businesses. We also show how the technique can identify the most salient BSC performance measures that can help predict overall performance. This ranking process is validated by the real-world outcome for the companies in terms of whether they were resilient (remained in operation) or went bankrupt.

Thus, not only do we seek to contribute to the general multiple-criteria analysis literature, but we also seek to contribute to the performance management literature to show how the BSC can be used to evaluate organizational resilience. The approach we introduce uses a statistical model, latent class analysis within a Bayesian framework, to form ranked clusters and then determine

* Corresponding author. Tel.: +1 508 831 4831.

E-mail address: jsarkis@wpi.edu (J. Sarkis).

company ranks within those clusters. Monte-Carlo simulation is used to estimate the parameters of interest and to assess the goodness of fit of the model.

This paper has three major objectives that are each a contribution to the literature. We develop a multiple-criteria ranking analysis method that requires only a minimal amount of subjective, managerial input. We apply the technique to a performance evaluation using aggregate financial and non-financial measures. We also examine the extent to which different BSC perspective measures are predictive of e-commerce companies' future performance and resilience through a metric we have called the displacement index.

The remainder of this paper begins with a brief review of relevant literature on performance measurement tools in general. Then we introduce the integrative multi-stage performance analysis methodology that rank-orders the companies based on a set of performance objectives. At this stage, we concurrently use real-world data to describe the methodology in detail. This methodology includes simulation using a Gibb's sampler to extricate latent variable used to determine the cluster ranks and to assess the goodness of fit of the model. Next, we introduce a validity measure called the displacement index that is used to determine the ability of a BSC perspective to gauge future viability of a firm. Finally, we present some overview of the results, limitations, and potential extensions of the methodology.

2. Multiple-criteria analysis and performance measurement: methods and applications

The performance evaluation of companies, departments, or individuals (generically referred to as objects) is necessary for managerial operations and decision-making. Outside the fields of engineering and management, performance measurements are necessary for research and practice, and the need to measure performance has not been lost to other disciplines from the sciences to the humanities. The development of tools and models for those areas also has been a continuous research endeavor for decades (Koksalan et al., 2011). In this section, we briefly review performance measurement tools, discuss one particular application, the BSC technique in more detail. This review sets the stage for the next section, which introduces the latent class ranking model with an application to e-business resilience.

The variety of tools available for performance measurement has increased greatly as new algorithms, problem situations, and supporting technology have evolved. Various tools and techniques have been developed or refined over the years. A summary of multiple criteria evaluation tools is presented in Table 1. There are various tradeoffs associated with each approach, ranging from

level of decision maker effort, to inclusion as well as flexibility of inclusion of various factors.

Various types of multiple criteria evaluation classifications can be found in the literature (e.g., see Figueira et al., 2005; Wallenius et al., 2008). The classifications include the types of tools used, the level of involvement in the technique, or other characteristics.

The research on organizational performance has relied on many of the techniques mentioned in Table 1. Multiple dimensions have been used to determine strategic and operational success potential in organizations. For example, when seeking to identify or select suppliers the multiple performance dimensions are used (Sarkis and Talluri, 2002; Agarwal et al., 2011). Multiple-criteria performance evaluations have been developed for strategic purposes and applied for issues such as the long-term resilience and bankruptcy evaluation (Ravi Kumar and Ravi, 2007). Many of these models in bankruptcy evaluation have focused on financial dimensions. We believe that considering only financial dimensions in these evaluations may be shortsighted and not encompass many intangible factors that can prove to be better predictors of overall corporate resilience. This latter area of investigation is where we seek to apply our modeling technique.

Thus, utilizing a tool such as BSC, which has been applied to management performance evaluations, to determine the strategic resilience of an organization may be an important opportunity for managers and analysts. Extant research on BSC follows several major streams. Kaplan and Norton (1996) have put a great deal of emphasis on linkages of outcomes from lower level (operational) perspectives to higher-level performance drivers. Without such linkages, BSC reverts to stand-alone sets of performance measures. Several papers in this area try to establish such linkages using different types of multivariate regression analyses. A BSC model that has linkages spanning a greater strategic perspective helps explain a firm's revenues, costs, profits and total assets better than a model that only has linkages to the next higher (typically operational) perspective (Bryant et al., 2004).

Generally speaking the different perspective used in a BSC model do interact with each other and the result observed are compounded outcomes of more than one perspective (for example: the financial perspective in our example is impacted by customer service perspective). Studies by Tjader et al. (in press), Hsu et al. (2011) and Kaplan and Norton (2008) have helped clarify the structure of perspective interactions. According to the research, it is not possible to account completely and correctly for the perspective interaction without an overly complex model. However, in many practical situations the main effects of the perspectives are robust enough that a manager can glean useful information without considering the compounding interactions between perspectives.

Another stream of research has focused on whether measures that are common to all business units of a firm receive more weight in

Table 1
Summary of multiple criteria evaluation technique characteristics and exemplary references. H=high, M=medium, L=low.

Evaluation technique	Cost of implementation	Data requirements	Ease of sensitivity	Economic rigor	Decision maker involvement	Management understanding	Mathematical complexity	Parameter mixing–flexibility
AHP	M	M	L	L	H	M	L	H
DEA	M	M	L	M	L	L	H	M
Expert systems	H	H	L	H	M	M	H	H
Goal Program	M	M	M	H	M	L	H	L
MAUT	H	H	M	M	H	M	M	H
Outranking	M	M	L	M	H	L	M	M
Simulation	H	H	H	H	L	H	H	M
Scoring models	L	L	L	L	H	H	L	H
Latent class	H	L	L	M	L	L	H	H
Cluster model								

Adapted from Sarkis and Sundarraj (2000) except for the last row, the information for which is provided by the authors.

performance evaluation while those measures that convey unique aspects of a business unit's operations are ignored (Lipe and Salterio, 2000). Such a simplifying approach undermines the essence of BSC, which needs to have a scorecard aligned to a business unit's strategy.

One of the purported benefits of adopting BSC is eventual improved financial performance resulting from inclusion of non-financial measures in a strategic business plan. Firms that add these non-financial measures to a performance evaluation significantly improve their return on assets and stock market performance (Said et al., 2003).

The Bayesian approach to selection and ranking that we utilize in this paper uses a simple two-level Bayes model to select the best mean performance measure (Morris and Christiansen, 1996). It generates samples from the product normal posterior distribution of the means and obtains posterior probabilities that each of the means is the largest. The method we utilize falls within the sampling based techniques of the Bayesian approach (Morris and Christiansen, 1996; Goldstien and Spiegelhalter, 1996).

BSC and many performance methodologies do not provide guidance on how to “balance” the score i.e., how to select weights for different outcomes in a perspective to determine the overall organizational performance. This ambiguity has created serious

implementation problems. The subjectivity in assignment of the weights by evaluators, e.g., through AHP, has been identified as a major cause of failure of BSC and its eventual abandonment by a major financial services firm (Ittner et al., 2003). In the methodology we develop and employ here the weights are implicitly selected by the statistical model based on the similarity of the performance outcomes of different companies. Managerial subjectivity is minimized in our proposed approach, allowing for evaluation that is more objective.

Similar to tools such as data envelopment analysis (DEA), this technique requires little management input and subjectivity. Yet, the technique is capable of allowing for greater flexibility in the types of performance measures. Unlike DEA, our technique is not as data driven e.g., the relative performance scores in DEA can change greatly depending on the data sample. These and other issues facing our latent class methodology are revisited in Section 6.

3. A latent class analysis methodology for organizational evaluation

Latent class analysis is a novel method to categorize and rank objects into analogous clusters using objectively determined or

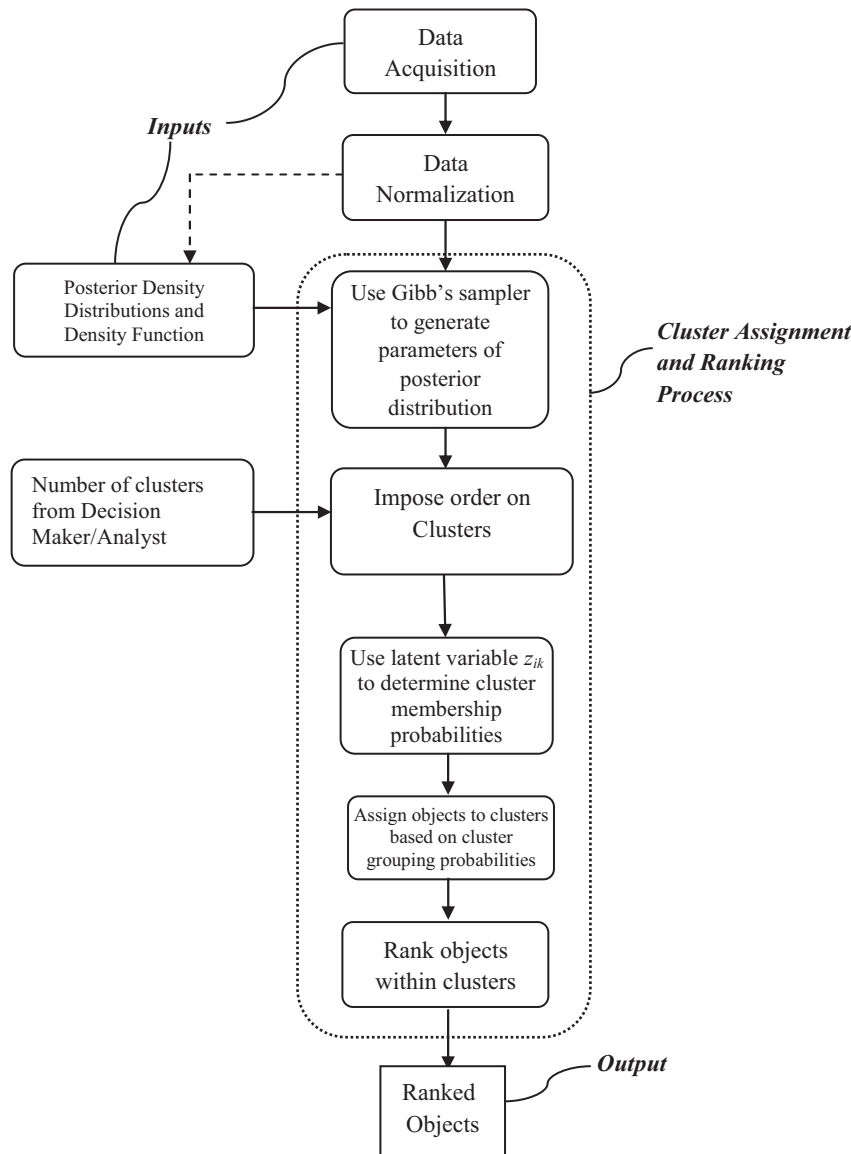


Fig. 1. Schematic diagram of the methodology.

observed characteristics of the objects. The ranking variable is unknown or latent and using a statistical technique its values can be determined. We summarize the methodology, which includes a number of steps and stages, in Fig. 1.

The methodology requires the following inputs: data about the objects' characteristics that will be used to rank the objects, the number of clusters the decision maker/analyst would like to have in the evaluation, and probability distribution function characteristics of various parameters. A first step is normalization of the data prior to the cluster assignment and ranking process. The first stage within the cluster and ranking process requires assimilation of the normalized data by a Gibbs sampler to determine the values of parameters by iterative simulation. These values are used to determine cluster membership probabilities for each object. Using membership probabilities, the objects are then assigned to clusters. The ranking process orders the clusters and then the objects within each of the clusters. The result is objects are ranked on an aggregated set of criteria.

We now introduce study data to illustrate the methodology as applied to BSC information for e-commerce (Internet) companies and determine the likelihood of their failure or success (resilience). The objects will be the companies while the criteria will be various BSC measures identified below. These data will serve as the input to the methodology, which we illustrate in succeeding sections.

3.1. Data collection

The following illustrative application of the methodology evaluates the resiliency of e-commerce companies based on BSC indicators. The goal is to determine whether and which indicators play a role in determining successful versus unsuccessful e-commerce organizations.

Unfortunately, the latent class cluster ranking method does not allow direct qualitative measurements. However, it does allow for ordinal numbers so if qualitative measurements can be transformed or are amenable to ordinal numbers it would be possible to include them.

The data are obtained from the e-commerce Almanac data set collected by the Intermarket Group. All 82 companies included in the Almanac were viable companies in the year 1999. However, only 42 of those that were considered "pure-play Internet companies" (i.e., those that use the Internet as their principal or sole sales channel) were possible candidates for the sample. Out of those companies, data on both financial and non-financial measures were available for only 20 companies. This final sample of companies represented a range of products and services that includes banking, financial services, books, music, prescription drugs, travel, online groceries, furniture, and automobiles. The reason for selecting the year 1999 is that it is the year just before dotcom bubble burst; several e-commerce firms went bankrupt in the following year. This business environment provides us with a rich source of data needed to test our model.

The specific data items selected for our analysis are number of unique visitors, revenue, income, marketing expenditure, development

expenditure, financing, number of employees, and the reach of a company. Our model does not use any proxy variables, which are common in many BSC studies. Table 2A provides definitions of variables used in this study.

In Table 2B we provide the data for the companies in our sample. The data have been 'standardized' by dividing the raw data for each matrix by the standard deviation of that metric for the 20 companies. Each row represents a company and each column represents a metric so each cell in a column is divided by the standard deviation of that column.

The data that are collected are for ecommerce startup firms only. It is very likely that the result derived in this paper may not be applicable to startups in other fields or to well-established companies due to the unique business model used and uncommon operating environment encountered in ecommerce startups.

3.2. Selection of financial and non-financial measures

The set of performance measures for e-commerce companies was selected from those listed by Olve et al. (1999). These performance measures, three each for the financial, customer, and business process perspectives of the BSC and two for the learning and growth perspective of BSC are shown in Table 3. The rationale for our choice of these performance measures is discussed briefly in the following sections.

3.2.1. Financial performance measures

Revenue growth, cost reduction, and asset utilization are three financial themes that typically drive business strategy (Kaplan and Norton, 1996). Thus, traditional financial measures include return on sales, return on assets, return on capital employed, revenue as a percentage of financing, and profitability. Since the financial robustness of a company is relevant to its ultimate survival, it was important to include some traditional measures in our analysis. Table 3 shows that under the financial perspective, we use the ratio of total revenue and financing, return on sales, and return on capital employed.

3.2.2. Customer measures

In Internet-based businesses, marketing expenditure and number of affiliates are used to generate visitors, some of whom will become customers and buy the products and services. Thus, revenues per unique visitors, marketing coverage, and market share (penetration) are used as performance measures to represent customer-relevant measures.

3.2.3. Business process measures

Each business has a unique set of processes for creating value for customers and producing financial results. For e-commerce companies, the number of employees and available financing influence the critical operations that must be performed to meet the needs of customers. Efficient use of these resources improves the conversion factor (i.e., numbers of unique visitors who become

Table 2A
Definition of data items collected.

Variable	Definition
Unique visitors	The number of visitors who visit a site more than once
Marketing expenditure	Amount used to attract prospects, which includes market research on target groups, sales support, mass advertising, etc.
Development expenditure	The expenditure on strategic investments geared towards innovation and differentiating performance, building strong customer loyalty, employee–systems–organizational alignment.
Financing	The total funds invested in a company from capital markets, including venture capital and initial public offerings.
Number of employees	The number of full time employees as of January 31, 2000.
Reach	The penetration level (reach) is obtained by assessing the proportion of users of a particular site to the entire profiled sites.

Table 2B
Standardized data.

	Financial			Customer			Internal business			L. & growth	
	(1) TR/F	(2) AP/TR	(3) AP/F	(4) TR/UV	(5) ME/UV	(6) MS	(7) TR/ME	(8) AP/N	(9) R/N	(10) EDC	(11) TR/DE
Amazon	2.367	0.001	0.006	1.082	0.943	4.400	1.505	1.095	0.004	1.22	1.039
Autobyl	1.098	0.302	1.174	0.394	1.493	0.305	0.346	0.909	1.176	3.679	0.286
Autoweb	1.218	0.373	1.61	0.146	0.510	0.671	0.375	0.763	1.222	1.358	0.651
Bolt, Inc.	0.300	2.801	2.974	0.037	0.265	0.345	0.184	0.131	1.577	1.202	0.127
CarsDirect	0.120	0.751	0.32	0.114	0.378	0.386	0.395	0.11	0.353	0.184	0.689
Cdnw	2.187	0.072	0.561	0.216	0.456	2.012	0.622	1.487	0.462	2.708	0.636
800.com	0.139	3.952	1.951	0.033	0.339	0.264	0.128	0.122	2.064	0.557	0.253
Drugstore.com	0.585	0.308	0.638	0.214	1.306	0.488	0.215	0.433	0.571	2.120	0.236
E-Loan Inc	0.342	0.515	0.625	0.410	1.942	0.163	0.277	0.320	0.708	0.597	0.621
eToys	0.160	0.319	0.181	0.253	0.605	0.305	0.548	0.162	0.221	0.223	0.840
E*Trade	1.291	0.019	0.089	2.468	4.143	0.732	0.781	1.314	0.110	1.860	0.818
Fogdog	0.186	1.726	1.136	0.068	0.721	0.305	0.124	0.259	1.916	1.462	0.205
FTD.com	3.916	0.245	3.400	0.491	0.410	0.305	1.568	3.355	3.527	1.672	2.324
Furniture.com	0.502	1.082	1.925	0.120	1.291	0.264	0.122	0.260	1.205	1.822	0.165
iOwn	0.964	0.795	2.716	0.444	1.989	0.102	0.293	0.260	0.887	2.094	0.144
NetB@nk	0.916	0.223	0.723	0.644	0.290	0.264	2.907	3.491	3.336	0.988	4.094
NextCard Inc	0.267	0.426	0.403	0.067	0.213	1.179	0.408	0.369	0.675	3.507	0.122
Peapod, Inc.	1.951	0.165	1.142	4.132	1.401	0.061	3.866	0.364	0.258	0.201	2.091
PlanetRx.com Inc	0.241	1.223	1.042	0.062	1.307	0.427	0.062	0.117	0.613	1.928	0.070
Webvan Group	0.053	0.771	0.146	0.929	2.836	0.041	0.429	0.068	0.223	0.885	0.088

Table 3
Performance measures used within each perspective.

Perspective	Strategic objective	Performance measure
Financial	Improve capital conversion Improve sales productivity Improve capital productivity	(1) Revenue/financing (2) Income/revenue (3) Income/financing
Customer	Improve revenue generation Engage in aggressive marketing Improve Market share (market penetration)	(4) Revenues/unique visitors (5) Marketing expenditure/unique visitors (6) Market share (percent of users captured by the firm)
Business process	Improve marketing effectiveness Improve employee efficiency Improve employee productivity	(7) Revenue/marketing expenditure (8) Income/number of employees (9) Revenue/number of employees
Learning and growth	Improve employee environment Spend development expenditure funds effectively	(10) Development expenditures/number of employees (11) Revenue/development expenditures

customers). Thus, revenues generated by per dollar of marketing expenditures, employee productivity, and visitor conversion rates by employee are included in the set of measures.

3.2.4. Learning and growth measures

In the face of the intense competition faced by e-commerce companies and the fact that the first mover advantage is lost relatively quickly, firms must make continual improvements and have the ability to introduce new and innovative products to continue to maintain their customer base. Employee skills, training, motivation, and empowerment are important in creating an environment in which innovation and product development and improvements are encouraged. Thus, development expenditure per employee, and revenue generated per dollar of development expenditure are included in the final set of measures.

3.3. Data normalization

A normalization process is initially completed to evaluate the acquired data. First, we define three index sets: $L=(1,2,...,i,...,l)$, the number of objects; $M=(1,2,...,j,...,m)$ the number of criteria; and $C=(1,2,...,k,...,c)$ the number of clusters. There are 20 companies (objects) and 11 performance measures in this case study data set ($l=20$, $m=11$). Let t_{ij} indicate the observed value of j th

performance measure for the i th company. Each value t_{ij} is divided by the standard deviation of that particular performance measure, $s_j=[\sum_L(t_{ij}-\bar{t}_j)/(l-1)]^{1/2}$, to make the observed values dimensionless and comparable. The dimensionless data are represented by $x_{ij}=t_{ij}/s_j$.

To disperse the data to approximate normality, we need to transform the data set. Three possible normalization transformations can provide one-to-one mappings and maintain a monotonically increasing transformation. The transformations will be either by cube root ($y_{ij}=\sqrt[3]{x_{ij}}$), square root ($y_{ij}=\sqrt{x_{ij}}$), or logarithmic ($y_{ij}=\ln(x_{ij})$).

To determine which transformation is best a Bayesian cross-validation analysis is completed by first obtaining deleted residuals on the dimensionless values for a given transformation n for every object i , and criterion j (DR_{ij}^n).¹ We then seek to determine the correlation between estimated values of each transformed data point y_{ij} and its corresponding deleted residual data point (DR_{ij}^n).² Estimates for each data point are determined using a Monte Carlo simulation with the Gibbs sampler (Gelfand and Smith, 1990).

¹ See Appendix A for calculation of deleted residual values.

² The reduced data set does not include the value for k , m over the full set of objects and criterion (see Appendix A).

Table 4
Best fit analysis for transformation and normalization of data.

Transformation	Correlations between the deleted residuals and predicted values
Logarithm	−0.192
Square root	−0.062
Cube root	−0.057

The transformational goal is to achieve ideally a zero correlation between the deleted residuals and the predicted reduced values and an ideal zero slope (which indicates a good fit). As evident from Table 4, rescaling the observed values by taking the cube root gives the best fit and an almost zero correlation (−0.057) between a deleted residuals prediction and the predicted values.

Given this validation result, we use the cube root transformed values of the performance measures y_{ij} , where $y_{ij} = \sqrt[3]{(t_{ij}/s_j)}$.

Rankings of n companies based on any one-performance measurement can result in $n!$ possible permutations. In our case, for the 20 e-commerce companies, we would have to compute about 2.4×10^{18} possible rankings. If two performance measures are considered simultaneously with ordered pair ranks, the possible number of rankings jumps to $(2.4 \times 10^{18})^2$. To avoid the large number of simulation iterations, we introduce the multi-step procedure for a ranking solution.

3.4. Determining posterior density distributions of variables

An important input to defining the latent variables (and other parameters) through a Gibbs sampler simulation, and eventually ranking the various companies, is the determination of posterior density distributions.

We have already defined how we arrive at the matrix \hat{y} denoted by, $\hat{y} = \{y_{ij} : \forall i \in L, j \in M\}$ consisting of transformed, normalized, values of 11 performance measures of 20 companies. Based on this known information, we would like to estimate the unknown parameters of interest such as $\hat{\theta}$, $\hat{\omega}$ and \hat{z}_i .³ Let all unknown parameters be denoted by the set $\Omega = \{\hat{\mu}, \hat{\nu}, \hat{\theta}, \hat{\omega}, \hat{z}, \hat{\sigma}\}$. To estimate the unknown parameters, we resort to Bayes' theorem, and compute the joint posterior density $\pi(\Omega|\hat{y})$. It can be shown that

$$\pi(\Omega|\hat{y}) \propto \pi(\hat{y}|\hat{\mu}, \hat{\nu}, \sigma_1^2) \pi(\hat{\mu}, \hat{z}|\hat{\theta}, \hat{\omega}, \sigma_2^2) \pi(\hat{\nu}|\sigma_2^2) \pi(\hat{\theta}) \pi(\hat{\omega}) \pi(\hat{\sigma}^2) \quad (1)$$

It is possible to obtain a mathematical expression for $\pi(\Omega|\hat{y})$ since the shapes of the distributions on the right-hand side of the proportional sign are known. The first term is our linear model density function and it has a weighted normal distribution, the same is true of the second and the third terms. The prior probability density functions of $\hat{\omega}$ and $\hat{\sigma}$ are assumed to be non-informative and each of their values is equally likely. Additional details are provided in Appendix A. Even though a mathematical expression is available for $\pi(\Omega|\hat{y})$, it does not have a recognizable or known form and thus analytical inferences about the parameter values cannot be made, which is the reason for the use of a Gibbs sampler simulation approach.

One of the more important and information-rich density functions is associated with μ_i . To provide some insight into the process we specify the assumptions associated with this density function.

3.4.1. The density function for μ_i

We have to specify the probabilistic characteristic of μ_i , which reflects the magnitude of a company performance measure as discerned by the BSC method. If a company definitely belongs to the k th cluster, then it is straightforward to specify the probability density function of μ_i for the latent class model as a normal distribution with mean equal to $\text{logit}(\theta_k^*)$ and variance equal to σ_3^2 or $N\{\text{logit}(\theta_k^*), \sigma_3^2\}$, however, there is no foreknowledge about the cluster to which the company may belong.⁴ To accommodate this uncertainty, we consider all normal distributions over all clusters, and use a probability vector $\hat{\omega}_i = [\omega_{ik}, k \in C]$ to determine the weighted normal distribution of μ_i as follows:

$$\pi(\mu_i, z_{ik}|\hat{\theta}, \hat{\omega}_i, \sigma_3^2) = \prod_{k=1}^C (\omega_{ik} N(\text{logit}(\theta_k^*), \sigma_3^2))^{z_{ik}} \quad (2)$$

3.5. Managerial input on number of clusters for grouping

In this next step, we make a decision about the number of clusters. We assume four clusters ($c=4$) for our case example. Four clusters are sufficient to separate the 20 companies grouping with approximately five companies in each cluster. We wish to balance the size of the clusters with the number of clusters. Too many clusters or too many members within a cluster can cause difficulty in model fitting. The number of clusters is the only external, subjective input from the decision maker we need for the model.

3.6. The cluster assignment and ranking process

The cluster assignment and ranking process for each of the objects (companies) will utilize the transformed data, posterior probability assumptions, and number of clusters, as defined in the previous section. This step of the process will form clusters of companies that are similar to each other based on defined criteria and then ranking the companies within each cluster based on the probability of belonging to that cluster. Thus, eventually, all companies will be ranked.

To complete this process of clustering, we (1) utilize a Gibbs sampler to generate a latent variable (and other parameter posterior distributions), (2) impose order on clusters, (3) determine cluster membership probabilities, (4) assign objects to clusters, and then (5) rank objects within clusters. These five steps are evident in Fig. 1 inside the dashed box.

We obtain the ranking of companies based on performance measures in each BSC perspective. For example, in the case of the financial perspective, our model combines the three financial performance measures (see Table 3) and provides a single ranking of 20 companies. We also combine all four perspectives together and consider all 11 performance measures to form the overall ranking of the 20 firms. In all we have five different rankings of the firms. The weights for combining the performance measures to obtain a ranking are determined implicitly by the model without subjective input.

3.6.1. Generating a latent class model and other parameters

We obtain the clusters by fitting a latent class model. The latent variable is the most important one we will generate for assignment of a company to a performance cluster. The model probabilistically identifies the hidden clusters through the latent variables. We use a Bayesian approach because it permits exact inference via the Gibbs sampler (Gelfand and Smith, 1990). Specifically we use the following linear model:

$$y_{ij} = \mu_i + \nu_j + e_{ij}, \quad \forall i \in L, \forall j \in M \quad (3)$$

³ For our notation, a $\hat{\cdot}$ over the notation, $\hat{\theta}$, for example, denotes a vector or a matrix, as the case may be, of variable θ_{ij} .

⁴ Logit function and its use in the model are described in later section.

where the transformed performance measure y_{ij} is a random variable that has three components: μ_i is the effect of object (company) i , ν_j is the effect of criterion (performance measure) j and a random noise factor e_{ij} . We make the usual assumption that all e_{ij} have independent and identical normal distributions with zero mean and variance σ_1^2 . The same holds for all ν_j except the variance is σ_2^2 ; μ_i has a density function as defined in Section 3.4.1.

To estimate the posterior distribution parameters, we use Markov chain Monte Carlo simulation using the Gibbs sampler (Gelfand and Smith, 1990). The Gibbs sampler generates samples that converge on the target distribution, i.e., the posterior distribution from which we derive inferences about the latent variables \hat{z} . To perform the Gibbs sampling, one needs conditional posterior density distributions for each parameter given all other parameters and \hat{y} . Let Ω_a indicate the set Ω excluding the parameter a . Then, it is possible to derive a conditional posterior density distribution $\pi(\mu_i|\Omega_{\mu_i}, \hat{y})$ for the parameter μ_i ; it happens to be a weighted normal distribution. Conditional posterior distributions for other parameters are as follows: $\hat{\theta}$ has weighted normal distributions, $\hat{\sigma}$ has gamma distributions, and $\hat{\omega}$ has Dirichlet distribution with latent variables as parameters. Additional details on these distributions are provided in Appendix A.

These distributions are derived mathematically. For the sake of brevity, only two conditional posterior distributions, without derivation, are shown here. Recall that Ω_a is the set Ω excluding parameter a . The conditional posterior density function for μ_i is as follows:

$$\pi(\mu_i|\Omega_{\mu_i}, \hat{y}) = \text{Normal}\left(\lambda \left[\frac{\sum_M (y_{ij} - \nu_j)}{m} \right] + (1-\lambda) \left(\sum_C z_{ik} \theta_k^* \right), (1-\lambda)\sigma_3^2\right) \quad (4)$$

where

$$\lambda = \frac{\sigma_3^2}{\sigma_3^2 + (\sigma_1^2/m)}$$

The conditional posterior density function for σ_1^2 is

$$\pi(\sigma_1^2|\Omega_{\sigma_1^2}, \hat{y}) = \text{Gamma}\left(\frac{lm+a}{2}, \frac{b + \sum_L \sum_M (y_{ij} - \mu_i - \nu_j)^2}{2}\right) \quad (5)$$

The other parameters for which the conditional posterior distributions that have been derived but not shown here are σ_2^2 , σ_3^2 , $\hat{\theta}$, $\hat{\omega}$, \hat{z} and $\hat{\nu}$.

We run the simulation to generate the necessary data using a ‘thinning’ approach. We discard or “burn in” 5000 iterates to let the stochastic process reach a steady state before collecting any data. Thereafter, every 75th observation is taken from the joint posterior density distribution to obtain a random sample of 1000 observations of variables of interest including \hat{z} . The Gibbs sampler produces a Markov chain so the consecutive iterates are correlated. To obtain an uncorrelated random sample, one strategy employed is to “thin out” the iterates, as we have done here.

3.6.2. Ordering of clusters

A company may belong to any one of the c clusters but we need to impose a known order on the clusters so when the eventual cluster membership is known we also know the cluster ranking. To accomplish this, we create a parameter θ , $0 \leq \theta \leq 1$. The interval $[0, 1]$ is divided into c smaller segments such that $0 = \theta_0 < \theta_1 < \theta_2 < \dots < \theta_{c-1} < \theta_c = 1$. Thus, the third segment, for example, would have end-points θ_2 and θ_3 . Let the midpoint of the k th segment be denoted by $\theta_k^* = (\theta_{k-1} + \theta_k)/2$. It is clear that $0 < \theta_1^* < \theta_2^* < \dots < \theta_k^* < \theta_{c-1}^* < \theta_c^* < 1$. By performing the logit transformation of a midpoint, the range of θ^* is dispersed over the complete real line, providing symmetry and approximate

normality.⁵ From the definitions of performance measures in Table 3, it is clear that the measures are designed such that the higher value of a performance measure indicates superior performance. It will become apparent later that it also means companies with superior performance will belong to higher numbered clusters; thus cluster q has better performing companies than in cluster $(q-1)$.

3.6.3. Use latent variables to determine cluster membership probabilities

The latent variable predicts if a company belongs to a cluster. Its name underscores the fact that the information, which determines cluster membership of the companies, already exists in the values of performance measures; however, this information has not been extracted. The latent variables will do so when they are inserted into the model. Let latent variable $z_{ik} = 1$ if the i th company belongs to k th cluster, otherwise $z_{ik} = 0$. The linear model assumes that an observed performance measure is a random variable.

This represents the uncertainty that is present in assigning companies to the clusters and that probabilities must be assigned to latent variables. Let ω_{ik} indicate the probability that i th company is in k th cluster or $Pr(z_{ik} = 1; z_{ik'} = 0, k' \neq k, k' \in C) = \omega_{ik}$. Since the company can belong to any of the k clusters, $\sum_{k=1}^c \omega_{ik} = 1$. It is assumed that the latent variables are independently distributed.

From the simulation results, we are principally interested in values of the latent variable z_{ik} . The value $z_{ik} = 1; z_{ik'} = 0, k' \neq k, (k \in C)$ indicates a company $i, (i \in L)$, belongs to a cluster $k, (k \in C)$. There are 1000 simulation sample values for each of the latent variables. Let us indicate the sample values for any z_{ik} by $z_{ik}^h, (h = 1, \dots, n = 1000)$.

Ideally, if the i th company belonged to the k th cluster then every simulation run should have $z_{ik}^h = 1, (\forall h \in N)$. Due to statistical ambiguity of cluster definition and sampling errors, however, not all simulation runs will show that occurrence. To determine the overall tendency we compute the estimated probability of company i being in cluster k .⁶

$$p_{ik} = \left(\sum_N z_{ik}^h \right) / n, \quad \forall i \in L, \quad (6)$$

Table 5 shows the probabilities of a company belonging to a given cluster after the simulation was completed for the customer measures perspective of the BSC. As we can see the first company, Amazon.com, has a 96.3% probability of belonging to cluster #4 and a 2.8% probability of belonging to cluster #3, based on customer measures.

3.6.4. Assign objects (companies) to clusters

In the previous steps, we identified c clusters and the estimated probability of the company belonging to each one of the clusters. The next step is a simple assignment of a company i to a cluster k . The company is assigned to the cluster in which it receives the highest probability.

$$p_{ik}^* = \text{Max}(p_{ik'}, k' \in C) \quad (7)$$

Based on expression (7), we see Amazon.com belongs to cluster #4 and PlanetRx.com Inc. belongs to cluster #2. Table 6 summarizes the cluster assignments for customer measure perspective for each e-business in our data set.

⁵ Logit transformation of θ_k^* is $\text{Logit}(\theta_k^*) = \ln(\theta_k^*/(1-\theta_k^*))$. Values of logit function are negative for $\theta_k^* < 0.5$, and positive for $\theta_k^* > 0.5$.

⁶ Errors in estimating the probability of belonging to cluster are judged by numerical standard error, $\text{NSE} = \sqrt{p_{ik}(1-p_{ik})/n}$. Smaller NSE values indicate better estimates.

Table 5

Probabilities of cluster membership for each E-business based on customer-measures BSC perspective.

	Company	C#1	C#2	C#3	C#4
1	Amazon.com	0.000	0.009	0.028	0.963
2	Autobyel.com Inc	0.005	0.417	0.577	0.001
3	Autoweb.com	0.008	0.419	0.572	0.001
4	Bolt Inc.	0.037	0.499	0.133	0.000
5	CarsDirect.com Inc.	0.027	0.597	0.375	0.001
6	Cdnow.com Inc	0.000	0.219	0.701	0.080
7	800.com Inc	0.556	0.372	0.072	0.000
8	Drugstore.com Inc	0.009	0.413	0.575	0.003
9	E-Loan Inc	0.020	0.523	0.456	0.001
10	eToys Inc	0.014	0.496	0.488	0.002
11	E*Trade Group Inc	0.000	0.028	0.110	0.862
12	Fogdog Inc	0.166	0.622	0.212	0.000
13	FTD.com	0.003	0.385	0.609	0.003
14	Furniture.com Inc	0.076	0.589	0.335	0.000
15	iOwn	0.030	0.578	0.390	0.002
16	NetB@nk	0.002	0.331	0.646	0.021
17	NextCard Inc	0.017	0.465	0.517	0.001
18	Peapod Inc	0.002	0.328	0.637	0.033
19	PlanetRx.com Inc	0.120	0.617	0.263	0.000
20	Webvan Group Inc	0.041	0.614	0.345	0.000

Table 6

Cluster assignment for customer-measure metric.

	Company	Cluster
1	Amazon.com	4
2	Autobyel.com Inc	3
3	Autoweb.com	3
4	Bolt Inc	2
5	CarsDirect.com Inc	2
6	CDnow Inc	3
7	800.com Inc	1
8	drugstore.com Inc	3
9	E-Loan Inc	2
10	eToys Inc	2
11	E*TRADE Group Inc	4
12	Fogdog Inc	2
13	FTD.com	3
14	Furniture.com Inc	2
15	iOwn	2
16	NetB@nk	3
17	Nextcard Inc	3
18	Peapod Inc	3
19	PlanetRx.com Inc	2
20	Webvan Group Inc	2

3.6.5. Determine the rank of an object within a cluster

Once the companies are assigned to the clusters, they must be ranked within the clusters. To perform this within-cluster ranking, we use the stochastic ordering principle, which stipulates that between two stochastic processes, the process that has the higher probability of producing the same random quantity be ranked higher. Once the cluster assignment of the companies are known, the following cumulative probability for a company i which is assigned to cluster k is computed.

$$q_{ik} = \sum_{s=k}^c p_{is} \quad (8)$$

The cumulative probability indicates the likelihood of a company being assigned to the current cluster, k , or higher clusters, $k+1$ through c . The companies in any cluster except the first are ranked based on their cumulative probability value: the higher the value, the higher is the ranking in the cluster. The cumulative probability for every company in the first cluster is one, thus the ordering rule cannot be applied. Therefore, the companies in the

Table 7

Within a cluster and overall rankings for customer-measures metric.

Cluster 4 ranks: (the highest cluster)

- (1) Amazon.com (0.963) [1]
- (2) E*Trade Group Inc (0.862) [2]

Cluster 3 ranks

- (1) Cdnnow.com Inc (0.781) [3]
- (2) Peapod Inc (0.670) [4]
- (3) NetB@nk (0.667) [5]
- (4) FTD.com (0.612) [6]
- (5) Drugstore.com Inc (0.578) [7]
- (6) Autobyel.com Inc (0.578) [8]
- (7) Autoweb.com (0.573) [9]
- (8) Nextcard Inc (0.518) [10]

Cluster 2 ranks

- (1) eToys Inc (0.986) [11]
- (2) E-Loan Inc (0.980) [12]
- (3) CarsDirect.com Inc (0.973) [13]
- (4) iOwn (0.970) [14]
- (5) Webvan Group Inc (0.959) [15]
- (6) Furniture.com Inc (0.924) [16]
- (7) PlanetRx.com Inc (0.880) [17]
- (8) Fogdog Inc (0.834) [18]
- (9) Bolt Inc (0.632) [19]

Cluster 1 ranks

- (1) 800.com Inc (0.556) [20]

first cluster are ranked based on the probability of belonging to first cluster, p_{i1} ; the higher the probability, the lower the rank of the company. Now, the rankings of the companies within the clusters are known, the ranking of the clusters themselves is determined by $\hat{\theta}^*$ values; a higher θ^* for a cluster means a higher cluster rank.

For the customer-measures perspective BSC measurement, the e-Business company rankings are summarized in Table 7. The number in the parenthesis next to a company's name in Table 7 is the cumulative probability of belonging to that or a higher cluster. The number in the square bracket is the overall rank of the company, across the clusters, for the customer-measures criterion.

In this fashion, the rankings for the companies are determined based on each criterion separately. The criteria utilized were each of the perspectives by itself and all of the perspectives together. The model ranks the companies by applying the specified set of performance measures without explicitly using weights. Rankings provided by the model are therefore robust because they are not subject to researcher intervention other than specification of the number of clusters to be used.

Table 8 provides the overall and specific measures rankings for each of the companies. We will revisit these values in a later section to investigate potential relationships. First, we provide a goodness of fit evaluation of our latent class model.

3.7. Goodness of fit evaluation

We complete an additional simulation experiment to evaluate the goodness-of-fit for the clustering approach. For this simulation, we specify a value for parameter θ , which indirectly specifies the cluster number to which resulting simulated observations will belong. Let us call this the "correct" cluster. The latent class model cannot always assign the observed or simulated values to the correct clusters because of the probabilistic nature of the decision sequence. Let us call the cluster assigned by the latent class method the "assigned" cluster. In the goodness-of-fit evaluation,

Table 8
Rankings of E-commerce companies under overall and specific measures.

Company	Overall rank based on four perspectives together	Rank on financial measures	Rank on customer measures	Rank on business process measures	Rank on learning and growth measures
FTD.com	<i>1*</i>	<i>1*</i>	<i>6*</i>	<i>1*</i>	<i>2*</i>
NetBank	<i>2*</i>	17	<i>5*</i>	<i>2*</i>	<i>1*</i>
PeaPod	<i>3*</i>	<i>2*</i>	<i>4*</i>	18	19
E*Trade	<i>4*</i>	12	<i>2*</i>	<i>10*</i>	<i>5*</i>
CDnow	<i>5*</i>	18	<i>3*</i>	<i>5*</i>	<i>3*</i>
Amazon	<i>6*</i>	8	<i>1*</i>	15	<i>7*</i>
Autobytel	<i>7*</i>	10	<i>8*</i>	<i>3*</i>	<i>4*</i>
iOwn	<i>8*</i>	<i>3*</i>	14	<i>11*</i>	13
Autoweb	<i>9*</i>	<i>7*</i>	<i>9*</i>	<i>4*</i>	<i>8*</i>
Bolt	10	<i>4*</i>	19	<i>8*</i>	9
Webvan	11	9	15	16	20
PlanetRx	12	16	17	17	11
Nextcard	13	11	<i>10*</i>	<i>12*</i>	<i>6*</i>
Furniture	14	<i>6*</i>	16	<i>9*</i>	12
Fogdog	15	19	18	<i>6*</i>	17
eToys	16	15	11	19	18
E-Loan	17	14	12	<i>14*</i>	14
Drugstore.com	18	13	<i>7*</i>	<i>13*</i>	16
800.com	19	<i>5*</i>	20	<i>7*</i>	15
CarsDirect	20	20	13	20	10

Better performing companies, which belong to the top two clusters, are numbered in italics with an asterisk.

Table 9
Classification efficiency of the model probability of the model assigning the correct cluster in a simulation.

Correct cluster	Assigned clusters			
	1	2	3	4
1	85.34%	14.32%	0.33%	0.00%
2	9.16	56.75	33.54	0.55
3	3.06	41.68	52.70	2.56
4	0.00	0.07	10.80	89.13

we are trying to determine how often the method is able to assign clusters correctly. For this purpose, we generated 1000 data points with known “correct” clusters and let the latent class model classify the data points to “assigned” clusters. The conditional probabilities in Table 9 were computed based on the assignments by the latent class model to different clusters given the “correct cluster” (the first column of the table). The table shows the latent class model is very efficient in the classification of companies in the extreme clusters. The companies that belonged to the highest-ranked cluster have 89.13% probability, $P(C4|C4)$, of being correctly labeled. The probability of correct classification for the lowest-ranked cluster is 85.34%. The probability of correct placement in the middle two clusters is just over 50%. Care should be taken for rankings in middle groups since there can be significant room for inappropriate clustering and ranking.

4. Evaluation of overall results: financial versus non-financial measures

Having established the validity of the clusters as differentiating between the better and worse performing companies, the objective of the next step of the analysis is to determine if the specific non-financial perspectives that represent critical aspects of e-commerce business strategy (customer, and learning and growth) are associated with future firm performance. We now show how the results from Table 8 can be used to help investigate this question.

We can begin by investigating the concordance of the various measures (financial and non-financial) with the overall rankings. If non-financial measures are valuable for evaluating the performance of e-commerce companies, we can expect to find concordance between the overall rankings of the companies generated by taking all 11 measures together with the rankings based on the non-financial measures alone.

Table 8 presents the information about the ranking of the 20 companies on four different perspectives both separately, and simultaneously (overall) by evaluating all 11 performance measures together. The companies that belonged to clusters 4 or 3 are better performing companies. To differentiate them for lesser performing companies in clusters 2 or 1, the rank numbers of companies in the better performing clusters are shown in italics with an asterisk in the table. Note that the number of better performing companies in each column varies, corroborating the common perception that companies do not perform equally well in all dimensions of their operations.

The first column of the table presents companies ranked from the best to the worst based on their overall performance. The first nine companies fell in clusters 4 or 3, while the remaining 11 fell in cluster 2 or 1. The next four columns of the table show the rankings based on each individual perspective. It is apparent from the table that the ranks of the companies varied across financial and non-financial measures.

In order to examine the extent of concordance between the overall performance rank and the ranks on each perspective, Spearman's rank order correlations are determined in Table 10.

The results indicate that the rankings on the financial measures have the weakest correlation to the rankings of overall performance. The rankings on the business process perspectives were weakly correlated to the rankings of overall performance but were statistically significant at the 3% level. However, the rankings based on learning and growth and especially the customer perspectives were highly correlated with the rankings of overall performance. These results support our claim that non-financial measures, specifically customer, and learning and growth related measures, represent a match with the underlying strategic objectives that have been emphasized in the e-commerce industry.

Table 10

Spearman rank-order correlations between overall rank and rank on each BSC perspective.

	Rank on financial perspective	Rank on customer perspective	Rank on business process perspective	Rank on learning and growth perspective
Spearman coefficient	0.344 (n/s)	0.794***	0.444*	0.570**
Alpha error	0.137	≈ 0.000	0.050	0.009

n/s: statistically not significant at 3% level.

*** Statistically significant at 0.05% level in a one-tailed test.

** Statistically significant at 1% level in a one-tailed test.

* Statistically significant at 3% level in a one-tailed test.

Table 11

Overall rankings and subsequent history of E-commerce companies.

Company	Overall Rank based on four perspectives together	Continued viability through 2002	Merged or bought over 2000–2002	Bankrupt or closed 2000–2002
FTD.com	<i>1*</i>	Viable		
NetBank	<i>2*</i>	Viable		
PeaPod	<i>3*</i>		Purchased by Royal Ahold in June 2000	
E*Trade	<i>4*</i>	Viable		
CDnow	<i>5*</i>		Merged with Amazon 2002	
Amazon	<i>6*</i>	Viable		
Autobytel	<i>7*</i>	Viable		
iOwn	<i>8*</i>			Bankrupt—2001
Autoweb	<i>9*</i>		Merged with Autobytel 2001	
Bolt	10			Closed 2000
Webvan	11			Closed 2001
PlanetRx	12			Liquidated 2001
Nextcard	13			Bankrupt—2002
Furniture	14			Bankrupt in 2000, resurrected as part of Levitz 2002
Fogdog	15		Acquired by Global Sports 2000	
eToys	16			Bankrupt—2001
E-Loan	17	Viable		
Drugstore.com	18	Viable		
800.com	19		Bought over by Circuit City—2002	
CarsDirect	20			Closed 2000

Better performing companies, which belong to the top two clusters, are numbered in italics with an asterisk.

5. BSC measures and future firm performance

BSC calls for inclusion of external and internal measures of critical business processes that are considered the performance drivers of future firm performance (Kaplan and Norton, 1996). BSC can be transformed from just a measurement system into a strategic management system. It is in this context that the question of how the various perspectives can drive organizational resilience becomes particularly important. BSC includes performance measures and drivers that have a direct impact on a firm's survival but such drivers generally need a longer horizon to affect the firm's performance. Thus, assessments over time and linking of non-financial measures to future firm performance are critical to establish the utility of these measures.

Table 11 documents the subsequent history of each company in the two-year period following 1999. Each company's history was tracked by obtaining relevant announcements and financial information from publicly available databases. As shown in Table 11, eight out of the 11 companies in the lesser-performing clusters ceased to exist. Six were closed or declared bankruptcy, while two were bought out or acquired by brick and mortar companies. In contrast, only one of the nine companies in the better performing clusters declared bankruptcy, while two companies merged with other e-businesses.

The association between BSC measures, especially the non-financial measures that represent the e-commerce strategy, and future firm survivability of the companies are analyzed in two ways. First, we develop a measure called the displacement index

to determine the ability of different BSC perspectives to judge future viability based on the rankings determined in the latent analysis. Second, the association is examined by documenting the links between the rank orderings of the companies and their subsequent history of survival.

5.1. The displacement index

In this section, we determine the ability of BSC to predict the sample companies' resiliency. The rankings provided by the latent analysis model are used. If the perspective performance measures were able to judge the long-run financial well-being of the companies correctly, then it is logical to assume that they would rank the firms that survived the e-commerce bubble and remained viable higher than the ones that went bankrupt or otherwise ceased operations. The actual rank per se is not important in this analysis; what is important is whether a viable company belongs to the top segment of rankings and nonviable companies to the bottom segment of the rankings. An incorrect segment placement of a company reflects inability of a BSC perspective to gauge future continued existence.

Examination of the last column of Table 11 reveals that out of 20 companies in our sample, eight companies turned out to be nonviable at the end of the period. The companies are iOwn, Bolt, Inc., Webvan Group, Inc., PlanetRx.com, NextCard, Inc., Furniture.com, eToys, Inc., and Carsdirect.com, Inc. Ideally, if a BSC perspective (and its allied performance measures) has the perfect ability to judge future viability, then it would rank the 12 viable

Table 12

Calculation of displacement index for finance perspective.

Company	Rank, r from Table 8	R_k or R'_k	$(v+k-R_k)$ or $R'_k-(v-d+k)$ The least number of places a company must be moved to be in the correct segment
Top segment			
FTD.com	1	–	–
PeaPod	2	–	–
iOwn	3	$R_1 = 3$	$12 + 1 - 3 = 10$
Bolt	4	$R_2 = 4$	$12 + 2 - 4 = 10$
800.com	5	–	–
Furniture	6	$R_3 = 6$	$12 + 3 - 6 = 9$
Autoweb	7	–	–
Amazon	8	–	–
Webvan	9	$R_4 = 9$	$12 + 4 - 9 = 7$
Autobytel	10	–	–
NextCard	11	$R_5 = 11$	$12 + 5 - 11 = 6$
E*Trade	12	–	–
Bottom segment			
Drugstore.com	13	$R'_1 = 13$	$13 - (12 - 5 + 1) = 5$
E-Loan	14	$R'_2 = 14$	$14 - (12 - 5 + 2) = 5$
eToys	15	–	–
PlanetRx	16	–	–
NetBank	17	$R'_3 = 17$	$17 - (12 - 5 + 3) = 7$
CDNow	18	$R'_4 = 18$	$18 - (12 - 5 + 4) = 7$
FogDog	19	$R'_5 = 19$	$19 - (12 - 5 + 5) = 7$
CarsDirect	20	–	–
Total displacement	–	–	73
Displacement index	–	–	$73/20 = 3.15$

Non-viable companies are shown in italics.

companies in the first 12 places or in the top segment, and the eight nonviable companies would be ranked in the bottom segment, below the first 12.

It is important to underscore a couple of points relating to the displacement index analysis. The “correct” numerical ranking of the companies is a highly subjective concept because any ranking depends on the perspective that was set as the objective. The results in Table 8 shows that the rankings vary across the performance dimensions. For displacement index purposes, a perspective that is most efficient in gauging the future viability need not get the exact ranking “correct” within either the top or the bottom segment. Nevertheless, it should have all viable companies in the top segment, ranked in any order, in the range from one through 12, and the eight nonviable companies ranked 13–20 in the bottom segment. With a less efficient perspective, some viable companies may be ranked in the bottom segment and some nonviable ones in the top. When this happens, we measure the errors by the displacement index, which determines the efficacy of a perspective in judging future survival.

The other important point about the displacement index is that it is a not a relative measurement. For instance, the Spearman's rank correlation coefficients establish how closely the four individual perspectives follow the overall measure. The implicit assumption is that the overall measure is superior to individual measures, and indeed provides the “correct” ranking. If the overall measure fails to live up to this expected behavior, the correlation coefficient may not provide useful information. The displacement index measures the average minimum number of ranks misplaced without comparing to any “correct” benchmark rankings.

Let r indicate the rank of a company under a given perspective. Table 12 shows the displacement index calculations for the BSC finance perspective using information contained in Tables 8 and 11. Let n and v indicate the number of nonviable and viable companies; in our example, these values are eight and 12, respectively. Let d indicate the number of companies incorrectly placed in the top segment, then the number of companies

incorrectly placed in the bottom segment is also d . In Table 12, there are five companies in each segment. Let us define $R_k = r$, if and only if a company is the k th misplaced company in the top segment, where $k = 1, 2, \dots, d$. The variable R_k is not defined for companies that correctly belong in the top segment. The variable R'_k is defined similarly for the bottom segment. We assume all ranks are equidistant from each other; that is the distance between any two consecutive ranks is the same as any two other consecutive ranks. The total displacement for the top segment is computed as $\sum_{k=1}^d (v+k-R_k)$. This computation determines the least possible number of places misplaced companies in the top segment must be moved to correct the errors. The total displacement index for the misplaced companies in the bottom segment is given by $\sum_{k=1}^d (R'_k - (v-d+k))$. The fourth column in Table 12 displays these calculations. The displacement index is simply sum of the above two expressions divided by the number of companies involved, which in our case is $n+c$ equal to 20. The final expression can be simplified as follows:

$$\text{Displacement index} = \left(\left[\sum_{k=1}^d (R'_k - R_k) \right] + d^2 \right) / (n+c) \quad (9)$$

The value of the displacement index is zero when the companies that are viable end up in the top segment (without respect to the order) and those that are nonviable are in the bottom segment. When all companies that belong to one segment are incorrectly placed in another segment, the displacement index reaches its maximum value, which equals $\min\{n, c\}$. In our example, the displacement index can vary from 0 to 8. Table 12 shows the displacement values for all perspectives.

The results in Table 13 provide an interesting picture indicating which BSC perspectives are useful in predicting the future survivability of e-commerce companies. The common belief that the financial measures are the best predictors of future viability does not appear to hold true for the e-commerce case; in fact, they happen to be the worst predictors. The best predictor, the one with

Table 13

Displacement indices for different perspectives, displacement index here varies from 0 to 8. Smaller numbers indicate better prediction ability.

Perspectives	Displacement index average error in placement
Overall, all perspectives together	2.20
Financial measures	3.15
Customer measures	2.60
Business process measures	1.80
Learning and growth measures	2.90

the least number of errors in placing the companies in either viable or non-viable segments, is business process perspective. On an average, it misplaced companies only by 1.8 places. The financial perspective has the most misplacement errors; consequently, it is the least reliable in predicting future survivability according to the displacement index.

The performance measures used for business process perspective were marketing and employee effectiveness and productivity (see Table 3). It appears that the companies that were generally more efficient in delivery of goods and services and marketed them effectively had better chance of surviving in the e-commerce environment. Most e-commerce companies were in existence for just a few years by the year 1999 and they continued to face a rapidly changing business environment. One could make a case that the companies were still on a learning curve, and those companies (and the employees) that were quick to learn were able to survive.

The overall performance measure has the second lowest displacement error rate. One could hypothesize that it would be the best indicator of a company's resiliency since it takes into account all aspects of a company's operations. However, it is possible that the fast-changing business environment faced by e-commerce companies needed, most of all, greater internal efficiencies and highly adaptive business culture to survive the dot-com meltdown.

6. Concluding remarks

In this paper, we introduce a latent class model for ranking objects based on simultaneous consideration of multiple criteria and specifically focus on utilizing this technique for performance-based ranking of companies. The case companies are e-businesses, the performance criteria evolve from the balanced score card (BSC) dimensions. The technique utilizes a Gibbs sampler that uses a small data set and generates joint posterior distributions of latent variables within a Bayesian framework. These latent variables are used to rank various objects first in clusters then within each cluster. These rank orderings are used to determine the perspectives of a BSC method that are most closely aligned with the overall performance of a company, and to identify the perspectives that are able to predict the eventual survival of an e-commerce firm.

The use of latent class analysis in this paper avoids some significant issues inherent in other methodologies and in the BSC method itself. For example, the "Hawthorne Effect" (Ittner and Larcker, 1998a) is avoided by considering a sample of companies that had not implemented (to our knowledge) the balanced scorecard method. Ittner et al. (2003; 729) point to a "significant limitation" in many studies due to "the lack of data on non-financial or subjective performance dimensions, forcing researchers to use indirect proxies for the measures' informativeness". This study was able to use actual data on the 11 performance measures.

BSC also lacks a definitive procedure to determine the appropriate weights for outcomes and perspective for computation of performance measures. Obviously, the choice of weights significantly affects the performance measures. The model introduced in

this paper allows us to skirt the contentious task of assigning subjective weights to different perspectives, and to the measures within the perspectives (unlike in techniques such as AHP, out-ranking, and simple scoring). A major advantage of the latent class model is the latent variables can be part of a normal regression model, which makes the resulting classification and ranking procedure robust.

Our test of economic relevance, which examines whether BSC and its component perspectives provide valuable information about the long-run economic viability of companies, offers interesting results. For e-commerce companies, the financial measures were the least reliable in predicting whether a firm will continue to be viable in the future. The most reliable measure, as measured by the displacement index, was a company's performance on the business process perspective. This result may not be a surprise to those who believe in the nuts-and-bolts approach, and that attention to detail eventually pays off. The combined, overall performance measures are placed just behind the business process in their ability to predict the survivability of an e-commerce firm. Interestingly, the customer perspective, which ranked the highest in correlation with the overall performance, was not a strong predictor of a company's viability.

One of the practical and research results of this study underscore the concept that the financial perspective for e-commerce companies may not be as important as it is for the more established business environment. The study points to the specific non-financial measures—customer, and learning and growth perspectives—that represent critical aspects of the e-commerce business strategy as major determinants of an e-business's overall performance. Results of our study are generally consistent with some of the factors identified in popular books and periodicals on e-commerce as leading causes of the companies' success (May, 2000; Mellahi and Johnson, 2000; Seybold and Marshak, 2000). It should be noted that the results obtained in this paper are based on data collected from e-commerce startup firms. It is very likely that the results may not extend to startups in other fields or to any mature and stable companies.

The second set of results pertains to the connection between performance measures related to different perspectives and the future viability of the companies. This connection is developed using the displacement index based on the rank order obtained from the latent class model. We find that business process perspective is the best predictor of the future viability of an e-commerce company. The other non-financial perspectives also predict the viability with varying degrees of accuracy. Contrary to the prevailing expectation in other business environments, the financial metrics in e-commerce are not a good predictor of whether a company will survive in the long run.

Although we have introduced a novel ranking technique that we have shown can be used to generate practical and research insights, there are some limitations. These limitations clearly point to potential future developments. For example, there are issues with the 'overhead' of developing and implementing a simulation system to help generate the various assumed probability distributions. Knowing and identifying probability distributions is not always an easy process. Completing a sensitivity analysis of the results under different assumptions may be worthwhile, but will require significant computational effort. Development of a user-friendly decision support system may overcome this limitation, but the sensitivity of the final solution will need to be examined.

The technique also has some subjectivity in the determination of the number of clusters. We found that at higher and lower levels the clusters gave relatively consistent assignments of objects, but in the middle levels it was a bit more difficult to make assignments and determine the ranking. Investigating and identifying the optimal number of clusters may be needed before the application of the technique. Testing the number of clusters chosen, the size of

optimal sample, the number of criteria for ranking may provide some insights into the sensitivity the results of the model.

We have compared this technique qualitatively to other techniques, e.g., AHP, DEA, scoring. Future research could involve in-depth examination and evaluation using actual data to determine the performance of the methodology introduced in this paper with the other techniques. A difficulty would be to determine baseline analyses and may require practical experimentation with management. Also, each potentially comparative technique has their own nuances that would need to be evaluated in the comparison. For example, numerous DEA models and variations of AHP and other multiple criteria approaches exist. Careful study design would be necessary. It is very likely that the comparative approaches may be more complementary, rather than supplementary. Examination of complementary characteristics of the techniques for integrated multi-methodological developments with the latent class ranking with other approaches is a promising avenue for future research.

Finally, we feel that introducing the latent class ranking approach can set the stage for many other multiple criteria performance evaluation and decision applications. For example, complex decision environments such as sustainability, supplier management, location analysis and other operational and strategic decisions requiring simultaneous consideration of multiple criteria can benefit from the application of this technique. We feel that the number of potential applications and the insights gathered from these applications will be quite significant as researchers' creativity enhances the approach and applications.

Appendix A

Deleted residuals calculation

Deleted residuals determine which transformation will work best based on the correlation between predicted values from the original data and the transformed data. Let us define the set Y as the transformed set of data comprised of transformed elements y_{ij} . For example, the element $y_{ij} = \sqrt[3]{x_{ij}}$. Let us also define the transformed reduced set

$$Y_{(k,m)}' = \{y_{ij} | i \neq k, j \neq m, y_{ij} \in Y\}$$

Define $\mu_{k,m} = (\sum_{\substack{i \in L, i \neq k \\ j \in M, j \neq m}} y_{ij} / (l * m - 1))$ as the mean of the reduced transformed set.

$$\text{Define } \sigma_{k,m} = \left(\frac{1}{(l * m) - 2} \sum_{\substack{i \in L, i \neq k \\ j \in M, j \neq m}} (y_{ij} - \mu_{k,m})^2 \right)$$

The deleted residual for a transformation n , DR_{km}^n is now defined as

$$DR_{k,m} = \left(\frac{y_{k,m} - \mu_{k,m}}{\sigma_{k,m}} \right)$$

Density functions

Prior probability density functions

The probability vector $\hat{\omega}_i$ has the following prior density distribution:

$$\pi(\hat{\omega}_i) = \text{Dirichlet } (1, 1, 1, \dots, 1), \quad \text{i.e.,}$$

$$\pi(\hat{\omega}_i) = \begin{cases} \frac{1}{c} \sum_c \omega_{ik} = 1, & \omega_{ik} \geq 0. \\ \text{Otherwise} & \pi(\hat{\omega}_i) = 0 \end{cases}$$

The variances $\sigma_1^2, \sigma_2^2, \sigma_3^2$ have identical and independent Gamma prior distributions with parameters $a/2$ and $b/2$, $a=b=0.002$. This value of a and b is selected to provide a non-informative prior.

The prior probability distribution for $\hat{\theta}$ is

$$P(\hat{\theta}) = 1/c! \quad \text{when } 0 < \theta_1 < \dots < \theta_c,$$

$$\text{Otherwise } P(\hat{\theta}) = 0$$

References

- Agarwal, P., Sahai, M., Mishra, V., Bag, M., Singh, V., 2011. A review of multi-criteria techniques for supplier evaluation and selection. *Int. J. Ind. Eng. Comput.* 2, 801–810.
- Bititci, U.S., Suwignjo, P., Carrie, A.S., 2001. Strategy management through quantitative modelling of performance measurement systems. *Int. J. Prod. Econ.* 69 (1), 15–22.
- Bryant, L., Jones, D.A., Widener, S.K., 2004. Managing value creation within the firm: an examination of multiple performance measures. *J. Manage. Accounting Res.* 16, 107–131.
- Clivillé, V., Berrah, L., Mauris, G., 2007. Quantitative expression and aggregation of performance measurements based on the MACBETH multi-criteria method. *Int. J. Prod. Econ.* 105(1), 171–189.
- Figueira, J., Greco, S., Ehrgott, M. (Eds.), 2005. *Multiple Criteria Decision Analysis: State of the Art Surveys*. Springer, Berlin.
- Gelfand, A.E., Smith, A.F.M., 1990. Sampling based approaches to calculate marginal densities. *J. Am. Stat. Assoc.* 85, 398–409.
- Goldstien, H., Spiegelhalter, A., 1996. League tables and their limitations: statistical issues in comparisons of institutional performance. *J. R. Stat. Soc. (A)* 159, 385–443.
- Hsu, C.W., Hu, A.H., Chiou, C.Y., Chen, T.C., 2011. Using the FDM and ANP to construct a sustainability balanced scorecard for the semiconductor industry. *Expert Syst. Appl.* 38 (10), 12891–12899.
- Ittner, C.D., Larcker, D.F., 1998a. Innovations in performance measurement: trends and research implications. *J. Manage. Accounting Res.* 10, 205–238.
- Ittner, C.D., Larcker, D.F., Meyer, M.W., 2003. Subjectivity and weighting of performance measures: evidence from balanced scorecard. *Accounting Rev.* 78 (3), 725–758.
- Kaplan, R.S., Norton, D.P., 1992. The balanced scorecard—measures that drive performance. *Harv. Bus. Rev.* 70, 71–79.
- Kaplan, R.S., Norton, D.P., 1996. *The Balanced Scorecard: Translating Strategy into Action*. Harvard Business School Press, Boston, MA.
- Kaplan, R.S., Norton, D.P., 2008. *The Execution Premium: Linking Strategy to Operations for Competitive Advantage*. Harvard Business School Press, Boston MA.
- Koksalan, M., Wallenius, J., Zionts, S., 2011. *Multiple Criteria Decision Making—From Early History to the 21st Century*. World Scientific Publishing Co. Pte. Ltd., Singapore.
- Lipe, M.G., Salterio, S.E., 2000. The balanced scorecard: judgmental effects of common and unique performance measures. *Accounting Rev.* 75 (3), 283–298.
- May, P., 2000. *The Business of E-commerce: From Corporate Strategy to Technology*. Cambridge University Press, UK.
- Mellahi, K., Johnson, M., 2000. Does it pay to be a first mover in e-commerce? The case of Amazon.com. *Manage. Decision* 38 (7), 445–452.
- Morris, C.N., Christiansen, C.L., 1996. Hierarchical models for ranking and identifying extremes, with applications. In: Bernardo, J.M., Berger, J.O., David, A.P., Smith, A.F.M. (Eds.), *Bayesian Statistics*. Oxford University Press, Cambridge, United Kingdom, pp. 227–296.
- Olve, N., Roy, J., Wetter, M., 1999. *Performance Drivers: A Practical Guide to Using the Balanced Scorecard*. John Wiley and Sons, New York.
- Ravi Kumar, P., Ravi, V., 2007. Bankruptcy prediction in banks and firms via statistical and intelligent techniques—a review. *Eur. J. Oper. Res.* 180 (1), 1–28.
- Said, A.A., HassabElnaby, H.R., Weir, B., 2003. An empirical investigation of the performance consequences of non-financial measures. *J. Manage. Accounting Res.* 15, 193–224.
- Sarkis, J., Talluri, S., 2002. A model for strategic supplier selection. *J. Supply Chain Manage.* 38 (1), 18–28.
- Sarkis, J., 2003. Quantitative models for performance measurement systems—alternate considerations. *Int. J. Prod. Econ.* 86 (1), 81–90.
- Sarkis, Joseph, Sundarraj, R.P., 2000. Factors for strategic evaluation of enterprise information technologies. *Int. J. Phys. Distrib. Logistics Manage.* 30 (3–4), 196–220.
- Seybold, P.B., Marshak, R.T., 2000. *Customers.Com: How to Create a Profitable Business Strategy for the Internet and Beyond*. Random House, New York, NY.
- Suwignjo, P., Bititci, U.S., Carrie, A.S., 2000. Quantitative models for performance measurement system. *Int. J. Prod. Econ.* 64 (1–3), 231–241.
- Tjader, Y., May, J.H., Shang, J., Vargas, L.G., Gao, N., 2013. Firm-level outsourcing decision making: a balanced scorecard-based analytic network process model. *Int. J. Prod. Econ.* (<http://dx.doi.org/10.1016/j.ijpe.2013.04.017>, in press.).
- Wallenius, J., Dyer, J.S., Fishburn, P.C., Steuer, R.E., Zionts, S., Deb, K., 2008. Multiple criteria decision making, multi-attribute utility theory: recent accomplishments and what lies ahead. *Manage. Sci.* 54, 1336–1349.